A unified framework for evaluating bounds on the Bayesian cost

Article *in* Signal Processing · December 1998

DOI: 10.1016/S0165-1684(98)00151-0 · Source: DBLP

Wasfi Kafri
Birzeit University

ELSEVIER

Signal Processing 71 (1998) 279–289

SIGNAL PROCESSING

A unified framework for evaluating bounds on the Bayesian cost

W.A. Hashlamoun[a,*], W.S. Kafri[b]

[a] Department of Electrical Engineering, Birzeit University, P.O. Box 14, Birzeit, West Bank, Palestine, Israel

[b] Department of Electrical Engineering, Birzeit University, P.O. Box 14, Birzeit, West Bank, Palestine, Israel

Received 13 March 1998; received in revised form 13 August 1998

Abstract

A unified framework for evaluating bounds on the average cost of an optimum Bayesian receiver with arbitrary cost assignments is presented. The framework is developed based on formulating the binary hypothesis testing problem from a decision-theoretic perspective. This formulation results in a representation for the minimum average cost that is analogous to that for the minimum probability of error. Taking advantage of this analogy, a whole new series of generalized bounds on the minimum average cost is obtained by employing the well-developed theory of bounds of the minimum probability of error problem available in the literature. To demonstrate the applicability of the proposed unified framework, two upper bounds on the minimum cost, that generalize the known Bhattacharyya and Chernoff upper bounds on the minimum probability of error, are derived. The unified framework is also used to obtain a new generalized class of upper and lower bounds in terms of a modified form of the *f*-divergence. All new bounds derived in the paper are shown to reduce to the probability of error bounds under special cost assignments. © 1998 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Ein einheitliches System zur abschätzung der Schranken des Risiko eines Bayesschen optimalen Empfänger mit willkürlich gegebenen Kosten ist gebracht. Das System brüht auf die binäre Testhypothese der Entscheidungstheorie. Diese Formulierung bringt Analogie der Darstellung des Risikos mit der minimalen Fehlerwharscheinlichkeit. Dadurch ist die wohlentwickelte Theorie der minimalen Fehlerwahrscheinlichkeit benutzt, um eine Reihe verallgemeinerten Schranken des Risikos zu erreichen. Um die Anwendbarkeit des einheitlichen Systems zu demonstrieren, zwei obere Schranken des Risikos sind abgeleitet, die die bekannten Bhattachayya und Chernoff obere Schranken der minimalen Fehlerwahrscheinlichkeit verallgemeinern. Das einheitliche System is auch benutzt, um eine allgemeine Klasse von oberen und unteren Shranken in Beziehung der *f*-divergenz abzuleiten. Unter spezifische Kostenvorgabe es ist gezeigt, daß alle neu abgeleitete Schranken in dieser Arbeit die Schranken der Fehlerwharscheinlichkeit reduzieren. © 1998 Elsevier Science B.V. All rights reserved.

Résumé

Cet article présente un cadre unifié pour l'évaluation des bornes sur le coût moyen d'un récepteur Bayésien optimal avec assignation arbitraire du coût. Le cadre est basé sur la formulation de l'hypothèse de test binaire dans une perspective basée sur la théorie de la décision. Cette formulation donne naissance à une représentation du coût minimum

* Corresponding author. Tel.: 9722 298 2112; fax: 9722 298 2984; e-mail: hwael@eng.birzeit.edu.

PII: S0165-1684(98)00151-0

moyen qui est analogue à celle de la probabilité minimum d'erreur. Profitant de cette analogie, toute une série de bornes généralisées est obtenue en utilisant la théorie des bornes de la probabilité minimum d'erreur. Pour démontrer l'applicabilité du cadre unifié proposé, nous avons développé deux bornes supérieures du coût minimum, généralisant les bornes bien-connues de Bhattacharyya et Chernoff. Le cadre unifié est également utilisé pour l'obtention d'une nouvelle classe de bornes supérieures et inférieures d'une forme modifiée de la f-divergence. Nous démonstrons que toutes les nouvelles bornes proposées dans cet article réduisent les bornes de la probabilité d'erreur en utilisant des assignations de coût particulières.

Keywords: Bayesian cost; Probability of error; f-divergence; Chernoff bound; Bhattacharyya bound; Convex functions

1. Introduction

The probability of error decision rule is used when one has to make a decision in a binary hypothesis testing problem in those cases where the a priori probabilities of the hypotheses are known or at least can be found in a reliable way. However, in many problems of practical importance the optimum decision rule can be derived but an exact evaluation of the probability of error is very difficult or even impossible. For these cases, bounds on the error probabilities or approximate expressions for these probabilities that are easier to evaluate should be developed. Upper and lower bounds based on approximating the function $\min(p,1-p)$, which appears in the Bayes' error function, by well-behaved functions have been proposed. The use of these bounds as tools for the design and performance assessment of optimum receivers is based on the intuition that the tighter is the bound the closer is the performance of the receiver to the optimum one.

The problem of bounding the minimum probability of error has received a considerable amount of interest and a number of upper and lower bounds on the minimum probability of error have been proposed in the literature, e.g. [1–5,9,10,12,13]. Because of the strong link between the probability of error and the class of Ali–Silvey distance measurers (or the f-divergence), a number of the bounds available are expressed based on these distance measures. The idea behind this comes from Blackwell's Theorem [12] which states the following. Let π be the set of all permissible pairs of the a priori probabilities, π_0 and π_1, in a binary hypothesis testing problem. Then, there exists a subset of π for which if the distance between a given set of conditional density functions is larger than the distance between another set of conditional density functions, then the probability of error corresponding to the first set is less than the probability of error corresponding to the second set. Boekee and Van der Lubbe [3] provided an upper bound on the probability of error in terms of the f-divergence between the conditional density functions under the two hypotheses. A lower bound on the probability of error in terms of the f-divergence is provided in [2]. The Bhattacharyya bound [12], expressed in terms of the Bhattacharyya coefficient, is relatively simple to evaluate and has closed form expressions for many commonly used distributions. The bound has been used as a criterion for the design of quantizers for hypothesis testing [15] and for the design of distributed detection systems [14]. The Chernoff bound [10] provides an upper bound on the probability of error in terms of a scalar $s, 0 \leqslant s \leqslant 1$. The tightest bound is obtained by optimizing the upper bound with respect to the scalar s. A tighter bound than the Bhattacharyya bound is in terms of the equivocation function [5]. The Bayesian bound introduced by Devijver [7] is known to be tighter than both the equivocation and the Bhattacharyya bounds. The sine-Gaussian bound introduced in [8] is tighter than all the above mentioned bounds. Other bounds can be found in [1,4]. However, almost all of the literature in the area of bounds has been limited to finding bounds on the minimum probability of error. Little [7,8] can be found on the general Bayesian problem in which arbitrary costs are assigned to each course of action in the decision process.

In this paper, we reconsider the Bayesian receiver with arbitrary cost assignments and derive an alternative representation for the minimum average cost of such a receiver. This representation is then used to obtain new generalized bounds on the minimum cost using existing approximation functions and techniques bounding the minimum probability of error such as the Chernoff bound, the Bhattacharyya bounds, and bounds in terms of the f-divergence. The rest of the paper is organized as follows. In Section 2, we present a general framework for establishing bounds on the minimum probability of error. In Section 3, we extend the results of Section 2 to the general Bayesian problem where we present a unified framework for establishing bounds on the minimum average cost. As a direct application of this unified framework, we derive in Section 4 two generalized bounds on the minimum cost: the Bhattacharyya and the Chernoff Bounds. In Section 5, we employ the same framework to obtain a new class of bounds on the minimum cost in terms of a modified form of the f-divergence. Section 6 is a discussion and conclusion.

2. Bounds on the minimum probability of error

Consider the binary hypothesis testing problem to test hypothesis H_0 against hypothesis H_1 with associated a priori probabilities π_0 and π_1. The decision is made based on a random observation X with conditional probability density functions $f_0(x)$ and $f_1(x)$ when H_0 and H_1 are true, respectively. The average probability of error is given by

$$P_{\mathrm{a}}(E) = \pi_0 P_{\mathrm{F}} + \pi_1(1 - P_{\mathrm{D}}), \tag{1}$$

where P_{F} is the probability of deciding H_1 when H_0 is true (probability of false alarm) and P_{D} is the probability of deciding H_1 when H_1 is true (probability of detection). The average probability of error in Eq. (1) is minimized when the likelihood ratio test is employed [17]. The optimum receiver is known as the maximum a posteriori probability (MAP) receiver. When an observation x is received, the MAP receiver computes the a posteriori probabilities $P(\mathrm{H}_0|x)$ and $P(\mathrm{H}_1|x)$ as

$$P(\mathrm{H}_j|x) = \frac{\pi_j f_j(x)}{f_X(x)}, \quad j = 0,1, \tag{2}$$

where $f_X(x) = (\pi_0 f_0(x) + \pi_1 f_1(x))$ is the unconditional probability density function of the random observation X. Then, it decides in favor of the hypothesis with the larger a posteriori probability. The conditional probability of error [10] is

$$\begin{aligned} P(E|x) &= \min(P(\mathrm{H}_0|x), P(\mathrm{H}_1|x)) \\ &= \min(p, 1-p) =: r(p), \end{aligned} \tag{3}$$

where

$$p =: P(\mathrm{H}_0|x), \quad 0 \leqslant p \leqslant 1.$$

The minimum probability of error, denoted by $P(E)$, is the expected value of $P(E|x)$ with respect to x, i.e.,

$$P(E) = \int_x r(p) f_X(x)\,\mathrm{d}x. \tag{4}$$

This expression for the minimum probability of error is exact but is computationally undesirable in many applications due to the discontinuity in the derivative of the function $r(p)$ at the point $p = 0.5$. Many of the upper and lower bounds available in the literature on $P(E)$ are derived based on functions of p, $r^*(p)$, with computationally desirable properties that can approximate the triangular function $r(p)$. The conditions imposed on the approximation functions are discussed in [9]. When a function $r^*(p)$ is substituted for $r(p)$ in Eq. (4), it will provide a close approximation for $P(E)$. Let $B_E(r^*)$ denote a bound on $P(E)$, then

$$B_E(r^*) = \int_x r^*(p) f_X(x)\,\mathrm{d}x. \tag{5}$$

The closer is the function $r^*(p)$ to $r(p)$, the tighter is the bound. Several existing bounds on $P(E)$ proposed in the literature can be interpreted using the above framework. For example, the function $r^*(p)$ for some important cases is given by

1. Bhattacharyya bound [12]

$$r^*(p) = \sqrt{p(1-p)}. \tag{6a}$$

2. Equivocation bound [5]

$$r^*(p) = -0.5p\log p - 0.5(1-p)\log(1-p). \tag{6b}$$

3. Bayesian bound [7]

$$r^*(p) = 2p(1-p). \tag{6c}$$

4. Sine–Gaussian bound [9]

$$r^*(p) = 0.5(\sin \pi p)\exp[-1.8063(p-0.5)^2]. \tag{6d}$$

These functions are plotted in [9] along with the function $r(p)$.

In the next section, we extend the above formulation of the probability of error problem to the general Bayesian problem where we present a unified framework for evaluating bounds on the minimum average cost.

3. Bounds on the minimum average cost

Let us consider again the binary hypothesis testing problem of Section 2 to test hypothesis H_0 against hypothesis H_1. Let C_{ij}, $i,j = 0,1$, denote the cost of deciding H_i when H_j is true. Then, the average cost per decision made by the receiver is given by

$$R_a = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} P(H_j, H_i), \tag{7}$$

where $P(H_j,H_i)$ is the probability of the joint event that H_j is true and H_j is decided. It is well known [11,17] that the cost in Eq. (7) is minimized when the likelihood ratio test is used. This test partitions the observation space of X into two disjoint and mutually exclusive optimum decision regions

$$S_0 =: \{x: (C_{10} - C_{00})\pi_0 f_0(x) > (C_{01} - C_{11})\pi_1 f_1(x), \tag{8a}$$

$$S_1 =: \{x: (C_{10} - C_{00})\pi_0 f_0(x) < (C_{01} - C_{11})\pi_1 f_1(x), \tag{8b}$$

such that when x falls in S_k, H_k is declared true. It is straightforward to show that the minimum value of R_a, denoted in the sequel by R, can be determined from Eq. (7) using S_k as

$$R = C_{00}\pi_0 + C_{11}\pi_1 + \int_x \min(C_0\pi_0 f_0(x), C_1\pi_1 f_1(x))\,dx, \tag{9}$$

where

$$C_0 = C_{10} - C_{00},$$

$$C_1 = C_{01} - C_{11}$$

and the integration is performed over the entire observation space of X, i.e., over S_0 plus S_1. Eq. (9) is valid under the usual assumption [17] that making a wrong decision is more costly than making a correct one. This implies that C_0 and C_1 are positive since $C_{10} > C_{00}$ and $C_{01} > C_{11}$.

A more appropriate form of R can still be obtained from Eq. (9) by introducing the variable u as follows:

$$u = \frac{C_0\pi_0 f_0(x)}{\tilde{f}(x)}, \tag{10a}$$

where

$$\tilde{f}(x) = C_0\pi_0 f_0(x) + C_1\pi_1 f_1(x).$$

Then, we have

$$1 - u = \frac{C_1\pi_1 f_1(x)}{\tilde{f}(x)}. \tag{10b}$$

Substituting Eqs. (10a) and (10b) into Eq. (9), we get

$$R = C_{00}\pi_0 + C_{11}\pi_1 + \int_x r(u)\tilde{f}(x)\,dx, \tag{11}$$

This representation constitutes the basis of our unified framework to develop generalized bounds on R in this paper. The domain of the variable u in (10a) is the interval $0 \leqslant u \leqslant 1$, which is the same as the domain of the variable p in Eq. (3). In addition, the function $r(u)$ in Eq. (11) is the same as the function $r(p)$ in Eq. (4), in which p has simply been replaced by the dummy variable u. By looking back at the expressions for $f_X(x)$ and $\tilde{f}(x)$, we conclude from the above comparison that the third term on the right-hand side of Eq. (11) has a functional behavior similar to that on the right-side of Eq. (4). This implies, except for a constant term, that the representation for R in Eq. (11) is analogous to that for $P(E)$ in Eq. (4). But there has been a considerable amount of work done on bounding $P(E)$ in Eq. (4), resulting in a well-developed theory of bounds. Therefore, with a slight modification, this whole work can now be directed towards establishing generalized bounds on R. By taking advantage of the analogy between Eqs. (11) and (4), we are now able to derive a whole new series of upper and lower bounds on R using available approximation functions and techniques bounding $P(E)$. If we let $r^*(u)$ be an approximation function of $r(u)$ and $B_R(r^*)$ be a bound on R, then by substituting $r^*(u)$ into Eq. (11) we get the following general bound:

$$B_R(r^*) = C_{00}\pi_0 + C_{11}\pi_1 + \int_x r^*(u)\tilde{f}(x)\,\mathrm{d}x. \qquad (12)$$

The function $r^*(u)$ in Eq. (12) could be any one of the functions given in Eq. (6a) or any other function that possesses the desired properties of an approximation function. For example, we may obtain a generalized equivocation bound by substituting Eq. (6b) into Eq. (12), or we may as well obtain a generalized Bayesian bound by using Eq. (6c) into Eq. (12). Two detailed examples on the utilization of Eq. (12) are given in Section 4. For each function $r^*(u)$ used, there will be a corresponding bound on R. The tightness of each bound depends directly on the degree by which $r^*(u)$ approximates $r(u)$.

4. Generalized Bhattacharyya and Chernoff upper bounds on R

To demonstrate the applicability of the unified framework presented above we derive, in Sections 4.1 and 4.2, two upper bounds on R that generalize the well-known Bhattacharyya and Chernoff upper bounds on $P(E)$.

4.1. Bhattacharyya upper bound on R

The bound we derive in this section will be expressed in terms of the Bhattacharyya coefficient ρ_B, given by

$$\rho_B = \int_x \sqrt{f_0(x) f_1(x)}\,\mathrm{d}x. \qquad (13)$$

The Bhattacharyya coefficient, which is easily seen to lie between 0 and 1, is a measure of the closeness of two density functions. For example, if $f_1(x) \equiv f_0(x)$, then $\rho_B = 1$, while if the support of $f_1(x)$ does not intersect with the support of $f_0(x)$, then $\rho_B = 0$.

To derive the generalized Bhattacharyya upper bound $\bar{R}_B$ on R, we start by using the approximation function Eq. (6a) into the general bound Eq. (12). Thus, we obtain

$$\bar{R}_B = C_{00}\pi_0 + C_{11}\pi_1 + \int_x \sqrt{u(1-u)}\tilde{f}(x)\,\mathrm{d}x. \qquad (14)$$

Substituting the values of u and $1-u$ from Eq. (10) into Eq. (14) and simplifying, we get

$$\bar{R}_B = C_{00}\pi_0 + C_{11}\pi_1 + \sqrt{C_0 C_1 \pi_0 \pi_1}\int_x \sqrt{f_0(x) f_1(x)}\,\mathrm{d}x. \qquad (15)$$

Substituting C_0 and C_1 from Eq. (9) into Eq. (15) and making use of the Bhattacharyya coefficient defined in Eq. (13), we get

$$\bar{R}_B = C_{00}\pi_0 + C_{11}\pi_1 + \sqrt{(C_{10}-C_{00})(C_{01}-C_{11})\pi_0\pi_1}\rho_B, \qquad (16)$$

which is the generalized Bhattacharyya upper bound on R.

4.2. Chernoff upper bound on R

The bound we derive next will be expressed in terms of the Chernoff coefficient ρ_C, given by

$$\rho_C = \min_{0\leqslant s\leqslant 1} \int_x (f_1(x))^s (f_0(x))^{1-s}\,\mathrm{d}x, \quad 0 \leqslant s \leqslant 1. \qquad (17)$$

To derive the generalized Chernoff upper bound $\bar{R}_C$ on R, we start with the following inequality which is true for any two positive real numbers a and b:

$$\min(a,b) \leqslant a^s b^{1-s}, \quad 0 \leqslant s \leqslant 1. \qquad (18)$$

Let $a = 1-u$ and $b = u$. Then, from Eq. (18) we have

$$\min(1-u,u) \leqslant (1-u)^s (u)^{1-s} = r^*(u), \qquad (19)$$

where $r^*(u)$ here is another candidate function that approximates $r(u)$. Substituting $r^*(u)$ on the right-hand side of Eq. (19) into Eq. (12), we get

$$\bar{R}_C = C_{00}\pi_0 + C_{11}\pi_1 + \int_x (C_1\pi_1 f_1(x))^s (C_0\pi_0 f_0(x))^{1-s}\,dx. \tag{20}$$

This upper bound is true for any value of s, $0 \leqslant s \leqslant 1$. The tightest bound, also denoted by $\bar{R}_C$ for notational simplicity, is obtained by minimizing the integral on the right-hand side of Eq. (20) with respect to s, i.e.,

$$\bar{R}_C = C_{00}\pi_0 + C_{11}\pi_1 + \min_{0\leqslant s\leqslant 1} \int_x (C_1\pi_1 f_1(x))^s (C_0\pi_0 f_0(x))^{1-s}\,dx. \tag{21}$$

Let s^* denote the particular value of s at which the minimum value is achieved. Then, by substituting C_0 and C_1 from Eq. (9) into Eq. (21) and making use of the Chernoff coefficient defined in Eq. (17), one obtains

$$\bar{R}_C = C_{00}\pi_0 + C_{11}\pi_1 + [C_{01} - C_{11})\pi_1]^{s^*}[(C_{10} - C_{11})\pi_0]^{1-s^*}\rho_C, \tag{22}$$

which is the desired generalized Chernoff upper bound on R.

Before we close this section we should mention that in the minimum probability of the error case, the bounds (16) and (22) reduce to the well-known Bhattacharyya and Chernoff bounds available in the literature [12,10].

5. A generalized class of bounds on the minimum average cost in terms of the modified f-divergence

The formulation of the Bayesian problem in Section 3 led to a representation for the minimum cost R that was similar, in functional behavior, to that for the minimum probability of error $P(E)$. This representation was employed in Section 4 to obtain generalized upper bounds on R in terms of the Bhattacharyya and the Chernoff coefficients. In this section, we will employ the same representation to derive a new class of generalized bounds on R in terms of a modified form of the f-divergence. The new bounds reduce to the minimum probability of error bounds derived in [3,2] under special cost assignments. We first consider, in Section 5.1, some properties of the f-divergence introduced by Csiszar [6]. We find it convenient, for the sake of the paper, to define a modified form of the f-divergence that takes into account the effect of the a priori probabilities and conditional costs on the distance between two conditional density functions in a general Bayesian problem. In Sections 5.2 and 5.3, respectively, we derive generalized upper and lower bounds on R in terms of the modified f-divergence. In Section 5.4, we specialize the results of the previous two sections to the Bhattacharyya-type bounds.

5.1. The modified f-divergence D and the minimum cost R

Let $f_0(x)$ and $f_1(x)$ be the conditional probability density functions of the random variable X under hypotheses H_0 and H_1, respectively. A measure of closeness of these two density functions can be expressed in terms of the f-divergence defined as the expected value under H_0 of a function of the likelihood ratio. It is given by [6]

$$D_f(f_0(x), f_1(x)) = \int_x h\left(\frac{f_1(x)}{f_0(x)}\right) f_0(x)\,dx, \tag{23}$$

where $h(u)$, $u \geqslant 0$, is a convex real-valued function of u that satisfies the following conditions:

$$\lim_{u\to 0} h(u) = h(0), \tag{24a}$$

$$0.h(0/0) = 0, \tag{24b}$$

$$0.h(c/0) = c \lim_{u\to\infty} h(u)/u, \quad 0 < c < \infty. \tag{24c}$$

Some of the examples of the f-divergence include the J-divergence ($h(u) = (u-1)\log u$), the Kullback–Leibler numbers ($h(u) = -\log u$ and $h(u) = u\log u$), the negative of the Matsusita–Hellinger–Bhattacharyya coefficient ($h(u) = -\sqrt{u}$), and the Kolmogorov variational distance ($h(u) = |u - 1|$).

For the sake of this analysis, we find it convenient to define the modified f-divergence as

$$D =: \int_x h\left(\frac{C_1 \pi_1 f_1(x)}{C_0 \pi_0 f_0(x)}\right) C_0 \pi_0 f_0(x)\, dx, \tag{25}$$

where C_0 and C_1 are the constants given in Eq. (9). Note that this distance measure, which is a generalization of Eq. (23), incorporates the a priori probabilities, π_0 and π_1, and the conditional costs, C_{ij}, $i, j = 0,1$, in evaluating the closeness of the density functions in a general Bayesian problem. Dividing both the numerator and denominator of the argument of the function $h(\cdot)$ in Eq. (25) by $\tilde{f}(x)$ and recalling the definition of the variable u in Eq. (10a), we see that Eq. (25) can be expressed in terms of u as

$$D = \int_x u h\left(\frac{1-u}{u}\right) \tilde{f}(x)\, dx. \tag{26}$$

Now define the function $g(u)$ for $0 \leqslant u \leqslant 1$ as

$$g(u) =: u h\left(\frac{1-u}{u}\right), \quad 0 \leqslant u \leqslant 1. \tag{27}$$

It has been shown in [16] that $g(u)$ is a convex function over the interval $0 \leqslant u \leqslant 1$ when $h(\cdot)$ is a convex function in its argument. In this section, we restrict the evaluation of the bounds on R to a class of convex functions $g(u)$ that are symmetric about the point $u = 0.5$, i.e., functions that satisfy the symmetry condition $g(u) = g(1 - u)$. The non-symmetric case can be handled by slightly modifying the approach presented here, but this will not be pursued any further. In Fig. 1, we depict a typical symmetric convex function $g(u)$, $0 \leqslant u \leqslant 1$, bounded between two symmetric triangular functions $\bar{y}(u)$ and $\underline{y}(u)$. These two functions will be used in Sections 5.2 and 5.3 to obtain upper and lower bounds on R in terms of D. They are members of a family of functions described by the equation

$$y(u) = \begin{cases} -au + b, & 0 \leqslant u < 0.5, \\ -a(1-u) + b, & 0.5 \leqslant u \leqslant 1, \end{cases} \quad a > 0, \tag{28}$$

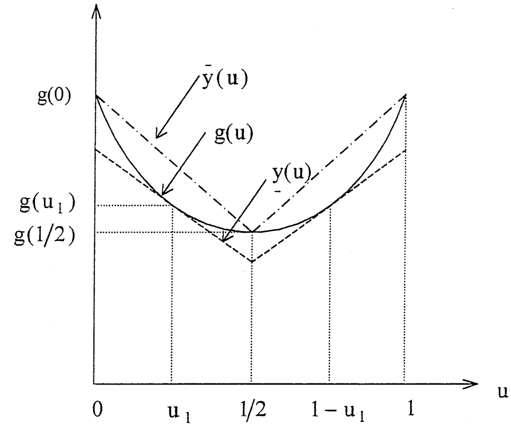

Fig. 1. The symmetric convex function $g(u)$ shown upper bounded by $\bar{y}(u)$ and lower bounded by $\underline{y}(u)$.

in which the parameters a and b assume specific values for both $\bar{y}(u)$ and $\underline{y}(u)$. Multiplying $y(u)$ in Eq. (28) by $\tilde{f}(x)$ and integrating over all x, we get

$$I = \int_x y(u) \tilde{f}(x)\, dx. \tag{29}$$

In Appendix A, we show that I is related to R by the following relationship:

$$I = b(C_0 \pi_0 + C_1 \pi_1) - a(R - C_{00} \pi_0 - C_{11} \pi_1). \tag{30}$$

From Fig. 1, we have the following bound on $g(u)$:

$$\underline{y}(u) \leqslant g(u) \leqslant \bar{y}(u), \tag{31}$$

in which equality of the upper bound holds at the points $u = 0$, $u = 0.5$ and $u = 1$, while equality of the lower bound holds at the two tangent points u_1 and u_2. Multiplying each term in this inequality by $\tilde{f}(x)$ and integrating over all values of x, we get the following bound:

$$\underline{I} \leqslant D \leqslant \bar{I}, \tag{32}$$

where $\underline{I}$ and $\bar{I}$ represent the value of I in Eq. (30) when $\underline{y}(u)$ and $\bar{y}(u)$ are, respectively, substituted for $y(u)$ in Eq. (29), and D is the modified f-divergence defined in Eq. (25). Since both $\underline{I}$ and $\bar{I}$ are functions of R, as evident from Eq. (30), then it is obvious that by a proper manipulation of Eq. (32), one can obtain the desired bounds on R in terms of D. This is the subject of the next two sections.

5.2. An upper bound on R in terms of D

In this section, we will make use of the results of Section 5.1 to obtain an upper bound on R in terms of D. We begin by referring to Fig. 1 in which we depict $g(u)$ upper bounded by $\bar{y}(u)$, a member of the family of functions $y(u)$ given in Eq. (28). The slope a can be easily evaluated from Fig. 1 as

$$a = \frac{g(1) - g(\frac{1}{2})}{1 - \frac{1}{2}} = \frac{h(0) - \frac{1}{2}h(1)}{\frac{1}{2}} = 2h(0) - h(1), \quad (33)$$

where $h(0)$ and $h(1)$ represent the value of $h(u)$ at $u = 0$ and $u = 1$, respectively. The second step in Eq. (33) follows from the relationship between $h(u)$ and $g(u)$ in Eq. (27) in addition to the properties of $h(u)$ in Eq. (24). The value of b can be evaluated from Eq. (28) and (33) as

$$b = \bar{y} + a(1 - u)|_{u=1} = g(1) = h(0). \quad (34)$$

By employing inequality (32), with the value of $\bar{I}$ obtained by using a from Eq. (33) and b from Eq. (34), into Eq. (30), we get

$$D \leqslant h(0)(C_0\pi_0 + C_1\pi_1) - (2h(0) - h(1))(R - C_{00}\pi_0 - C_{11}\pi_1). \quad (35)$$

Substituting C_0 and C_1 from Eq. (9) into Eq. (35), simplifying terms, and arranging for R on one side of the inequality, we get

$$R \leqslant \bar{R}, \quad (36)$$

where

$$\bar{R} = \frac{h(0)(C_{00}\pi_0 + C_{11}\pi_1 + C_{10}\pi_0 + C_{01}\pi_1) - h(1)(C_{00}\pi_0 + C_{11}\pi_1) - D}{2h(0) - h(1)}$$

is the desired upper bound on R. Note that $\bar{R}$ is given in terms of the value of the function $h(u)$ at the points $u = 0$ and $u = 1$, the a priori probabilities, the conditional costs, and the modified f-divergence D.

5.3. A lower bound on R in terms of D

Next, consider the function $\underline{y}(u)$ shown in Fig. 1 lower bounding $g(u)$. Its two segments are tangents to $g(u)$ at the points u_1 and u_2. Our objective in this section is to determine the parameters a, b, u_1 and u_2 so that $\underline{y}(u)$ provides the tightest lower bound on R. But these parameters are not completely independent. In fact, we will now see that they all can be expressed in terms of u_1. The derivative of $g(u)$ evaluated at the tangent point $u = u_1$ represents the slope of $\underline{y}(u)$ over the interval $0 \leqslant u < 0.5$. This means that

$$a = -g'(u)|_{u=u_1}. \quad (37)$$

Equating $g(u)$ to $\underline{y}(u)$ at the point $u = u_1$ results in

$$g(u_1) = -au_1 + b. \quad (38)$$

Substituting the value of a from Eq. (37) into Eq. (38) and solving for b, we get

$$b = g(u_1) - u_1 g'(u_1). \quad (39)$$

From Eqs. (37) and (39), we realize that both a and b are functions of u_1. Furthermore, from the symmetry of $\underline{y}(u)$ and $g(u)$ about $u = 0.5$, we realize also that $u_2 = 1 - u_1$. Hence, we have now all unknown parameters expressed in terms of u_1. By employing inequality (32) again, with the value of $\underline{I}$ obtained by substituting a from Eq. (37) and b from Eq. (39) into Eq. (30), we have

$$(g(u_1) - u_1 g'(u_1))(C_0\pi_0 + C_1\pi_1) + g'(u_1)(R - C_{00}\pi_0 - C_{11}\pi_1) \leqslant D. \quad (40)$$

Solving this inequality for R (note from Eq. (37) that $g'(u = u_1)$ is negative since a is positive), we obtain the following lower bound:

$$R \geqslant (C_{00}\pi_0 + C_{11}\pi_1) + u_1(C_0\pi_0 + C_1\pi_1) + \frac{D - g(u_1)(C_0\pi_0 + C_1\pi_1)}{g'(u_1)} =: \underline{R}. \quad (41)$$

Since $\underline{R}$ is a function of u_1, then this means that there exists an infinite number of lower bounds that satisfy Eq. (41). The tightest lower bound $\underline{R}^*$, which maximizes $\underline{R}$, can be obtained by differentiating $\underline{R}$ with respect to u_1 and setting the derivative equal to zero. The optimum value of u_1, denoted by u_1^*, is then the solution to the equation

$$D = g(u_1^*)(C_0\pi_0 + C_1\pi_1), \tag{42}$$

which yields the following tightest lower bound:

$$\underline{R}^* = (C_{00}\pi_0 + C_{11}\pi_1) + u_1^*(C_0\pi_0 + C_1\pi_1). \tag{43}$$

It can be shown, by evaluating the second derivative of $\underline{R}$ with respect to u_1 at the point u_1^*, that $\underline{R}^*$ is indeed the maximum value of $\underline{R}$. Substituting C_0 and C_1 from Eq. (9) into Eq. (43), we get

$$\underline{R}^* = (C_{00}\pi_0 + C_{11}\pi_1) + u_1^*((C_{10} - C_{00})\pi_0 + (C_{01} - C_{11})\pi_1), \tag{44}$$

which is the desired lower bound on R. This lower bound, just like the upper bound in Eq. (36), is expressed in terms of the a priori probabilities and the conditional costs. Its dependence on $h(u)$ appears in the form of the parameter u_1^*, which is the solution to Eq. (42).

In the minimum probability of error problem, i.e., when $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$, the upper bound $\bar{R}$ in Eq. (36) and the lower bound $\underline{R}^*$ in Eq. (44) reduce to the upper and lower bounds on $P(E)$ derived in [3,2].

5.4. Generalized Bhattacharyya upper and lower bounds on R

In this section, we specialize the results of Sections 5.2 and 5.3 to the Bhattacharyya-type bounds on R. Similar results were obtained in [8] using a different approach. For the Bhattacharyya-type bounds, the underlying convex function $h(u)$ is

$$h(u) = -\sqrt{u}. \tag{45}$$

By using this function into Eq. (27), we obtain the symmetric convex function

$$g(u) = -\sqrt{u(1-u)}. \tag{46}$$

The modified f-divergence, obtained by substituting Eq. (46) into Eq. (26) and using the value of u from Eq. (10a), is

$$D = -\sqrt{C_0C_1\pi_0\pi_1}\int_x \sqrt{f_0(x)f_1(x)}\,\mathrm{d}x = -\sqrt{C_0C_1\pi_0\pi_1}\rho_{\mathrm{B}}, \tag{47}$$

A generalized Bhattacharyya upper bound on R can be obtained by substituting $h(0) = 0, h(1) = -1$, and the value of D from Eq. (47) into Eq. (36). The result is given by Eq. (16).

To obtain a generalized Bhattacharyya lower bound on R we start by substituting the value of D from Eq. (47) and $g(u_1^*)$ from Eq. (46) into Eq. (42). The result is

$$-\sqrt{C_0C_1\pi_0\pi_1}\rho_B = -\sqrt{u_1^*(1-u_1^*)}(C_0\pi_0 + C_1\pi_1). \tag{48}$$

Solving this quadratic equation for u_1^* and taking the root that corresponds to the interval $0 \leqslant u < 0.5$, one obtains

$$u_1^* = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\frac{C_0C_1\pi_0\pi_1}{(C_0\pi_0 + C_1\pi_1)^2}\rho_{\mathrm{B}}}. \tag{49}$$

Using this value of u_1^* into Eq. (44) and simplifying, we get the following lower bound on R:

$$\underline{R}_{\mathrm{B}} = \frac{1}{2}(C_{00} + C_{10})\pi_0 + \frac{1}{2}(C_{01} + C_{11})\pi_1 - \frac{1}{2}\sqrt{[(C_{10} - C_{00})\pi_0 + (C_{01} - C_{11})\pi_1]^2 - 4(C_{10} - C_{00})(C_{01} - C_{11})\pi_0\pi_1\sigma_{\mathrm{B}}^2}. \tag{50}$$

6. Conclusion

A unified approach to evaluate bounds on the average cost of an optimum Bayesian receiver with arbitrary cost assignments was presented. This approach, which was based on a formulation of the Bayesian problem, led to a representation for the minimum average cost that is similar, in functional behavior, to that for the minimum probability of error. Taking advantage of this similarity, new bounds on the minimum cost were evaluated by employing the available approximation functions and techniques bounding the minimum probability of error. To demonstrate its applicability, the unified approach was used to obtain two upper bounds on R that generalize the known Bhattacharyya and Chernoff upper bounds of the minimum probability of error problem. It was also used to obtain bounds on R in terms of the modified f-divergence. This distance measure was defined to take into account the effect of the a priori probabilities and conditional costs on the distance between the conditional density functions in a general Bayesian problem. All bounds derived in this paper reduce to the probability of error bounds under special cost assignments.

Notation

$P(E)$	probability of error of an optimum receiver
H_0	null hypothesis
H_1	alternative hypothesis
π_0	a priori probability of H_0
π_1	a priori probability of H_1
X	random variable based on which a decision is made
x	value assumed by the random variable X
$P(E\|x)$	probability of error given an observation x
$f_0(x)$	probability density function of X when H_0 is true
$f_1(x)$	probability density function of X when H_1 is true
$f_X(x)$	unconditional probability density function of X
$\tilde{f}(x)$	a function defined in Eq. (10a)
$r(p)$	a symmetric triangular function defined in Eq. (3)
$r^*(p)$	a function that approximates $r(p)$
C_{ij}	cost of deciding H_i when H_j is true; $i,j=0,1$
C_0,C_1	quantities defined in Eq. (9)
R	average cost of an Optimum Bayesian receiver
S_0	optimum decision region corresponding to H_0
S_1	optimum decision region corresponding to H_1
ρ_{B}	Bhattacharyya coefficient
ρ_{C}	Chernoff coefficient
$\bar{R}_{\mathrm{B}}$	Bhattacharyya upper bound on R
$\underline{R}_{\mathrm{B}}$	Bhattacharyya lower bound on R
$\bar{R}_{\mathrm{C}}$	Chernoff upper bound on R
s^*	optimum value of s that minimizes $\bar{R}_{\mathrm{C}}$
u	variable defined in Eq. (10a)
$h(u)$	a convex function of u
$g(u)$	a symmetric convex function of u defined in Eq. (27)
D	the modified f-divergence defined in Eq. (25)
$\bar{y}(u)$	a symmetric triangular function of u used to upper bound $g(u)$
$\underline{y}(u)$	a symmetric triangular function of u used to lower bound $g(u)$
$\bar{R}$	upper bound on R in terms of D
u_1	point at which $\underline{y}(u)$ is tangent to $g(u)$
u_1^*	Optimum value of u_1 that maximizes the lower bound on R in terms of D
$\underline{R}$	a lower bound on R in terms of u_1 and D
$\underline{R}^*$	tightest lower bound on R in terms of D

Appendix A.

Consider the family of symmetric triangular functions $y(u)$ defined in Eq. (28). Two members of this family, denoted by $\bar{y}(u)$ and $\underline{y}(u)$, are depicted in Fig. 1. Multiplying $y(u)$ in Eq. (28) by $\tilde{f}(x)$ and integrating over all x, we get

$$I = \int_x y(u)\tilde{f}(x)\,\mathrm{d}x. \tag{A.1}$$

In what follows, we will verify that the value of this integral is linearly related to the minimum cost R. To that end, we split I into two parts. Each part corresponds to one segment of $y(u)$. For the first part, u ranges over the interval (0, 0.5) implying that $u < 1-u$. From Eqs. (10a) and (10b), this means that $C_0\pi_0 f_0(x) < C_1\pi_1 f_1(x)$. By virtue of Eqs. (8a) and (8b), we conclude that the integration over the first segment of $y(u)$ should be performed over S_1. For the second part, u ranges over (0.5, 1) implying that $1-u<u$. From Eqs. (10a) and (10b), this means that $C_1\pi_1 f_1(x) < C_0\pi_0 f_0(x)$ and the integration over the second segment of $y(u)$ should then be performed over

S_0. Evaluating Eq. (A.1) over the two segments of $y(u)$, we get

$$I = \int_{S_1} (-au + b)\tilde{f}(x)\,\mathrm{d}x + \int_{S_0} (-a(1-u) + b)\tilde{f}(x)\,\mathrm{d}x. \tag{A.2}$$

Taking b and $-a$ as common factors and combining the integrals over S_0 and S_1, we have

$$I = b\int_{S_1+S_0} \tilde{f}(x)\,\mathrm{d}x - a\left\{\int_{S_1} u\tilde{f}(x)\mathrm{d}x + \int_{S_0}(1-u)\tilde{f}(x)\,\mathrm{d}x\right\}. \tag{A.3}$$

This integral simplifies to

$$I = b(C_0\pi_0 + C_1\pi_1) - a\int_x \min(u,1-u)\tilde{f}(x)\,\mathrm{d}x. \tag{A.4}$$

The first term results from the fact that the area under a probability density function is 1, while the second from (8). The value of the integral in Eq. (A.4) can be obtained from Eq. (11) as

$$\int_x \min(u,1-u)\tilde{f}(x)\,\mathrm{d}x = R - (C_{00}\pi_0 + C_{11}\pi_1). \tag{A.5}$$

Using this result into (A.4), we get

$$\begin{aligned} I &= b(C_0\pi_0 + C_1\pi_1) - a(R - C_{00}\pi_0 - C_{11}\pi_1) \\ &= b(C_0\pi_0 + C_1\pi_1) + a(C_{00}\pi_0 + C_{11}\pi_1) - aR, \end{aligned} \tag{A.6}$$

which is the desired value of I. This result clearly indicates that I is a linear function of the minimum cost R. In Sections 5.2 and 5.3, we make use of this expression for I when deriving upper and lower bounds on R in terms of the modified f-divergence.

References

[1] M. Ben-Bassat, Use of distance measures, information measures, and error bounds in feature evaluation, in: P.R. Krishnaiah, L.N. Kanal (Eds), Handbook of Statistics, Vol. 2, Chapter 35, North-Holland, Amsterdam, The Netherlands, 1982, pp. 773–791.

[2] D.E. Boekee, J.C. Ruitenbeek, A class of lower bounds on the Bayesian probability of error, Inform. Sci. 25 (1981) 21–35.

[3] D.E. Boekee, J.C.A. Van der Lubbe, Some aspects of error bounds in feature selection, Pattern Recognition 11 (1979) 353–360.

[4] C.H. Chen, On information and distance measures, error bounds, and feature selection, Inform. Sci. 10 (1976) 159–173.

[5] C.H. Chen, Theoretical comparison of a class of feature selection criteria in pattern recognition, IEEE Trans. Comput. C-20 (September 1971) 1054–1056.

[6] I. Csiszar, Information-type measure of difference of probability distributions and indirect observations, Studia Scientarium Mathematicerium Hungarica 2 (1967) 299–318.

[7] P.A. Devijver, On a new class of bounds on Bayes risk in multi-hypothesis pattern recognition, IEEE Trans. Comput. C-23 (January 1974) 70–80.

[8] W.A. Hashlamoun, Bounds on the Bayesian cost in optimum receivers, Signal Processing 59 (1997) 275–284.

[9] W.A. Hashlamoun, P.K. Varshney, V.N.S. Samarasooriya, A tight upper bound on the Bayesian probability of error, IEEE Trans. Pattern Anal. Mach. Intell. 16 (February 1994) 220–224.

[10] M.E. Hellman, J. Raviv, Probability of error, equivocation, and the Chernoff bound, IEEE Tran. Inform. Theory IT-16 (4) (July 1979) 368–372.

[11] C.W. Helstrom, Elements of Signal Detection and Estimation, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[12] T. Kailath, The divergence and Bhattacharyya distance measures in signal selection, IEEE Trans. Commun. Technol. 15 (February 1967) 52–60.

[13] L. Kanal, Patterns in pattern recognition: 1968–1974, IEEE Trans. Inform. Theory 20 (1974) 697–722.

[14] M. Longo, T. Lookabuaph, R.M. Gray, Quantization for decentralized hypothesis testing under communication constraints, IEEE Trans. on Information Theory 36 (March 1990) 241–255.

[15] H.V. Poor, J.B. Thomas, Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems, IEEE Trans. Commun. COM-25 (September 1977) 893–900.

[16] I. Vajda, On the f-divergence and singularity of probability measures, Period. Math. Hungar. 2 (1972) 223–234.

[17] H.L. Van Trees, Detection, Estimation, and Modulation Theory, Vol. 1, Wiley, New York, 1968.